AD-A244 701

Defense Nuclear Agency
Alexandria, VA 22310-3398

(1)

DNA-TR-91-134

The Smart Data Manager: A User-Friendly Approach to Data Storage, Retrieval, and Manipulation

Dave Sargis
Barbara Trautlein
Daniel Higgins
JAYCOR
P.O. Box 85154
San Diego, CA 92186-5154

January 1992

Technical Report

CONTRACT No. DNA 001-90-C-0060

Approved for public release;
distribution is unlimited.

Destroy this report when it is no longer needed. Do not return to sender.

PLEASE NOTIFY THE DEFENSE NUCLEAR AGENCY, ATTN: CSTI, 6801 TELEGRAPH ROAD, ALEXANDRIA, VA 22310-3398, IF YOUR ADDRESS IS INCORRECT, IF YOU WISH IT DELETED FROM THE DISTRIBUTION LIST, OR IF THE ADDRESSEE IS NO LONGER EMPLOYED BY YOUR ORGANIZATION.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
920101

3. REPORT TYPE AND DATES COVERED
Technical 900801 - 910130

4. TITLE AND SUBTITLE
The Smart Data Manager: A User-Friendly Approach to Data Storage, Retrieval, and Manipulation

5. FUNDING NUMBERS
C-DNA 001-90-C-0060
PE - 62715H
PR - RJ
TA - RB
WU - DH600800

6. AUTHOR(S)
Dave Sargis, Barbara Trautlein, and Daniel Higgins

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
JAYCOR
P.O. Box 85154
San Diego, CA 92186-5154

8. PERFORMING ORGANIZATION REPORT NUMBER
J200-91-0019/2655

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Defense Nuclear Agency
6801 Telegraph Road
Alexandria, VA 22310-3398
TDTT/Gaudet

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
DNA-TR-91-134

11. SUPPLEMENTARY NOTES
This work was sponsored by the Defense Nuclear Agency under RDT&E RMC Code T 4662 D RJ RB 60080 TDTT 5700 A 25904D.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report summarizes the development and application of a Smart Data Manager (SDM). The SDM was developed as an aid to the DNA in its technical management and utilization of extensive classified data bases. The primary objective of the work was to establish the feasibility of using microcomputers and commercial software to facilitate storage, retrieval, and manipulation of classified underground test (UGT) data.

14. SUBJECT TERMS
Smart Data Manager
Underground Test Data Base
Hypertext Techniques
MATLAB Program

15. NUMBER OF PAGES
46

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

CLASSIFIED BY:

N/A since Unclassified

DECLASSIFY ON:
N/A since Unclassified

PREFACE

The authors would like to acknowledge the invaluable help of a number of DNA Test Directorate personnel including: Lt. Colonel John Gaudet, Commander Jim Gogolin, Captain Russell Daniel, and Mr. Juan Quintana.

CONVERSION TABLE

Conversion factors for U.S. Customary to metric (SI) units of measurement

Multiply ⟶ **By** ⟶ **To Get**

To Get ⟵ **By** ⟵ **Divide**

angstrom	1.000 000 X E -10	meter (m)
atmosphere (normal)	1.013 25 X E +2	kilo pascal (kPa)
bar	1.000 000 X E +2	kilo pascal (kPa)
barn	1.000 000 X E -28	meter^2 (m^2)
British thermal unit (thermochemical)	1.054 350 X E +3	joule (J)
calorie (thermochemical)	4.184 000	joule (J)
cal (thermochemical)/cm^2	4.184 000 X E -2	mega joule/m^2 (MJ/m^2)
curie	3.700 000 X E +1	*giga becquerel (GBq)
degree (angle)	1.745 329 X E -2	radian (rad)
degree Fahrenheit	$t_k = (t°f + 459.67)/1.8$	degree kelvin (K)
electron volt	1.602 19 X E -19	joule (J)
erg	1.000 000 X E -7	joule (J)
erg/second	1.000 000 X E -7	watt
foot	3.048 000 X E -1	meter (m)
foot-pound-force	1.355 818	joule (J)
gallon (U.S. liquid)	3.785 412 X E -3	meter^3 (m^3)
inch	2.540 000 X E -2	meter (m)
jerk	1.000 000 X E +9	joule (J)
joule/kilogram (J/kg)(radiation dose absorbed)	1.000 000	Gray (Gy)
kilotons	4.183	terajoules
kip (1000 lbf)	4.448 222 X E +3	newton (N)
kip/inch^2 (ksi)	6.894 757 X E +3	kilo pascal (kPa)
ktap	1.000 000 X E +2	newton-second/m^2 (N-s/m^2)
micron	1.000 000 X E -6	meter (m)
mil	2.540 000 X E -5	meter (m)
mile (international)	1.609 344 X E +3	meter (m)
ounce	2.834 952 X E -2	kilogram (kg)
pound-force (lbs avoirdupois)	4.448 222	newton (N)
pound-force inch	1.129 848 X E -1	newton-meter (N-m)
pound-force/inch	1.751 268 X E +2	newton/meter (N/m)
pound-force/foot^2	4.788 026 X E -2	kilo pascal (kPa)
pound-force/inch^2 (psi)	6.894 757	kilo pascal (kPa)
pound-mass (lbm avoirdupois)	4.535 924 X E -1	kilogram (kg)
pound-mass-foot^2 (moment of inertia)	4.214 011 X E -2	kilogram-meter^2 (kg•m^2)
pound-mass/foot^3	1.601 846 X E +1	kilogram/meter^3 (kg/m^3)
rad (radiation dose absorbed)	1.000 000 X E -2	**Gray (Gy)
roentgen	2.579 760 X E -4	coulomb/kilogram (C/kg)
shake	1.000 000 X E -8	second (s)
slug	1.459 390 X E +1	kilogram (kg)
torr (mm Hg, 0°C)	1.333 22 X E -1	kilo pascal (kPa)

* The becquerel (Bq) is the SI unit of radioactivity; 1 Bq = 1 event/s.

** The Gray (Gy) is the SI unit of absorbed radiation.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

SECTION 1

INTRODUCTION

1.1 THE GENERAL PROBLEM.

This report briefly summarizes the development and an initial application of a Smart Data Manager (SDM). The work was completed by JAYCOR to assist the Defense Nuclear Agency (DNA) in its technical management and utilization of large amounts of classified underground test (UGT) data. The primary objective was to determine the feasibility of using microcomputers and commercial software to facilitate the storage, retrieval, and manipulation of UGT data.

This is an age when the sheer volume of documentation tends to grow more rapidly than capabilities to assimilate and utilize such information. Any government or private organization can easily be overwhelmed by the difficult job of developing and maintaining useful data bases which encompass thousands of classified documents. Massive amounts of classified information can even lead to a "file-and-forget" approach to handling the continual influx of classified documents. In other cases, only a few individuals in an organization may be familiar with the scope and depth of the classified inventory, and efficient utilization of this inventory may be virtually impossible. Lastly, since the existence of crucial information may not be apparent to key individuals in the organization, the absence of user-friendly data bases can seriously impair organizational efficiency and productivity.

1.2 A PROPOSED SOLUTION.

Personal computers (PCs) and recently-available commercial software provide new opportunities to effectively manage a big data base on a desk top. Hypertext software is a high-speed automated way of searching for, accessing, and utilizing information. Hypertext directories and files are designed to allow the user to fly through vast quantities of information and quickly find those items of particular interest. The PC thus enables the

user to readily access information in a very intuitive manner and at the desired level of detail. Irrelevant or distracting data can be filtered out, so that the problem of having too much information is minimized or eliminated.

The computerized management and utilization of its classified document inventory would equip the DNA's Test Directorate with compact and powerful tools to monitor, evaluate, and guide UGT programs. For example, recent developments in compact disk technology (i.e., CD-ROM, CD-WORM, and CD-WMRM) have resulted in transportable diskettes, each of which is capable of storing about 250,000 pages of text. New removable hard disks are now also available which can store 320 megabytes) of information. The disk access times are very short, on the order of 10 ms. These disks can be stacked in sequences of seven each, so that a single PC can simultaneously access more than 2-billion bytes of information (or 500,000 full pages of text). Disk storage technology is advancing so rapidly that paper documents are likely to soon become much less desirable than "computer documents."

1.3 OUTLINE OF THE REPORT.

JAYCOR's approach to a SDM is described in Section 2. Hypertext techniques and a graphical user interface are employed on an IBM-compatible PC. The general problem of UGT data management is discussed in Section 3, while Section 4 deals with a specific application of the SDM to nuclear device output information (i.e., documents that cover outputs of the Disko Elm device). Section 4 also briefly summarizes operating procedures for the SDM. Section 5 describes lessons learned in this pilot project, and recommendations are put forward in Section 6.

SECTION 2

THE SMART DATA MANAGER

2.1 GENERAL DESCRIPTION.

The SDM was designed to be user-friendly and fast in accessing, retrieving, and manipulating stored data. It is menu driven with extensive on-line help menus. Text, tables, and figures are supported by the SDM. Moreover, built-in command links permit the SDM to access outside software packages; in particular, tabular data can be plotted via outside graphics programs.

2.2 HOW THE SMART DATA MANAGER WORKS.

The SDM has been developed in modular form to enable new software packages to be implemented as desired. The main commercial software packages in the SDM include the WINDOWS 3.0 graphical environment, the GUIDE 3.01 hypertext program, and the MATLAB program to provide plotting capabilities.

WINDOWS 3.0 is a "Macintosh-like" operating system with multi-tasking capabilities (multiple software packages can be open and available to perform tasks). Note that WINDOWS 3.0 manages and allocates memory above the DOS RAM limit of 640 kilobytes.

The GUIDE 3.01 hypertext program runs within the WINDOWS 3.0 environment so that it can produce and link files with text, tables, and graphical data. Computer documents can be structured to fit the available information and be made interactive to enable data to be plotted and manipulated. The SDM has the features which are necessary to create interactive computer documents and to generate a corresponding data base.

GUIDE 3.01 relies on buttons to indicate "live" areas of a computer document. When a button is activated by a mouse, new information is brought forth and displayed by the

computer. Buttons can be used to open a document, to reveal additional levels of detail, to display related information such as footnotes or glossary items in a "pop-up" menu, and to access cross-referenced information in a different part of the document or even in another document. Other buttons can be used to simultaneously open several programs such as plotting routines, calculational programs, and spreadsheets.

GUIDE 3.01 computer documents can be structured and utilized in a unique and time-saving way.

The 286 PC version of MATLAB has also been incorporated into the SDM, primarily to provide high-quality plotting capabilities. MATLAB is not a WINDOWS application program, and must be accessed through DOS commands within the SDM. However, the SDM is programmed in modular form so that as new WINDOWS-compatible graphics programs (such as EXCEL 3.0, PIXIE, EASY PLOT, etc.) become available, they can be used to replace MATLAB.

MATLAB was selected as the plotting program for the SDM since no satisfactory WINDOWS graphics programs were initially available. Furthermore, MATLAB has excellent technical plotting and digitizing features along with a robust library of mathematical, statistical, and engineering methods which can be brought to bear on a variety of technical problems.

Two IBM-compatible PCs were employed in this project: a 33-MHz 386 PC in Albuquerque, and a 286 PC in San Diego. The 386 has 8 megabytes of extended memory, and the 286 has a 3-megabyte extended memory board. Since the documents of interest to DNA are classified, external hard drives for both machines consisted of 44-megabyte, SYQUEST drives with removable hard disks. A Hewlett Packard (HP) ScanJet+ scanner was used to import text, tables, and figures into the SDM, and the Omnipage software package was used for optical character recognition (OCR) and conversion of files to text (ASCII) and word processing formats. Word Perfect 5.1 was the word processor of choice and it was also useful in preparing tabular data for plotting programs. Word Perfect and Omnipage were

both used to identify and correct scanning errors. Various "paint" programs provided the capability to rotate "landscape" figures into "portrait" form (to improve the on-screen readability of computer documents).

Figure 1 is a diagram that shows how the SDM treats archival data sets. Paper documents with text and figures are optically scanned and some text files are converted into ASCII files via the Omnipage OCR package. (Of course, formatted information on computer diskettes could be directly entered into the SDM.) Text and tabular data in ASCII files require less memory by a factor of ten or more than the corresponding scanned pictures or Tagged Image File Format (TIFF) files. In addition, data in ASCII files can be readily manipulated to produce plots.

All information from paper documents is stored on removable hard disks (or, possibly, on optical disks) which feed external drives. The removable disks with classified files can then be protected like classified documents.

The GUIDE 3.01 software accesses and links pertinent information as it is brought into the SDM. Guide reference buttons are generated to link related items of information within a specific document or group of documents. This facilitates rapid searches and the quick retrieval of textual and graphical material. Expansion buttons enable the user to access increasingly detailed information on a specific topic. Pop-up or note buttons indicate footnotes or user-generated memos on specific items, while command buttons provide links to outside codes (for example, software packages for generating plots of tabular data). The different types of GUIDE buttons are shown in Figure 2. Note that the GUIDE 3.01 software can also generate extensive on-line help menus.

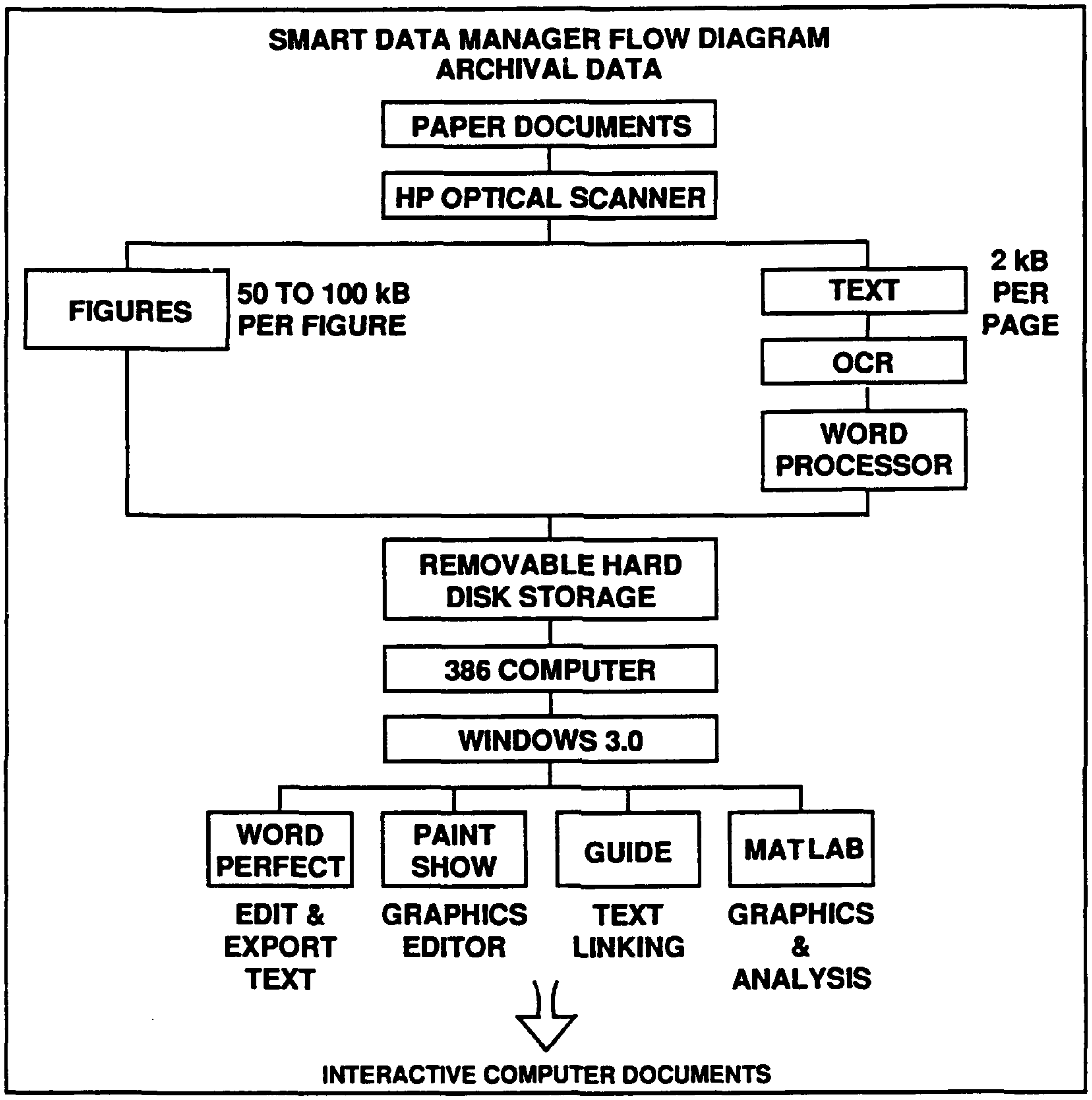


Figure 1. Smart Data Manager flow diagram for treating archival data sets.

[⇒*COMMAND BUTTONS*⇒]

[⊕**EXPANSION BUTTONS** ⊕]

[* <u>NOTE BUTTONS</u> *]

[⇨***REFERENCE BUTTONS***⇨]

Figure 2. Types of buttons and corresponding symbols used in GUIDE.

In addition to the above-mentioned software programs and capabilities, the SDM allows the user to indirectly access outside codes and data via a modem. For example, the MCNP Monte Carlo radiation transport program at Los Alamos has been accessed from within the SDM by simply clicking on the WINDOWS modem icon. A short MCNP run was executed and the generated data was returned via modem to the PC at JAYCOR for further manipulation. For longer calculations, the outside codes can be started, the results reviewed interactively, and data can be returned to the SDM at some later or more convenient time.

SECTION 3

UNDERGROUND TEST DATA PROBLEMS

Defense Nuclear Agency personnel have to deal with very large and complex data sets which span thousands of classified documents. There are two general types of UGT data sets of interest to the DNA, archival and real-time. The archival information covers dozens of UGTs that were performed over the last three decades. Real-time UGT data sets may be assembled to plan and manage a new UGT program or to rapidly evaluate a recent event.

There is a continuing need to maintain and update a base of information on nuclear device outputs, including measurements, predictions, and consensus results. Meaningful comparisons of measurements and predictions are always necessary. The DNA requires a special data base that facilitates such comparisons and that provides the user with some analytical capabilities.

Both archival and real-time data sets tend to be of a classified nature. A library of numerous classified documents is expensive to maintain and cumbersome to use. Furthermore, the existence and/or locations of pertinent classified information may only be known to a few individuals and the timely dissemination of this material to needy personnel can be difficult.

Another problem, common to all organizations, is the retirement or transfer of key technical personnel and managers. In-house experience and expertise can be seriously diminished when this takes place, and replacements are faced with an immense array of information and documentation which was assembled over many years. Needless to say, user-friendly data bases and computer documents can ameliorate this problem.

In planning and managing new UGT programs, numerous classified and unclassified documents are generated. These documents include proposals, memos, letters, data sheets, financial reports, status reports, schedules, drawings, calculations, etc.. Such a data base

must also be continuously updated as test beds and experiments become better and better defined. The computerized management of this type of information warrants further consideration by the DNA.

SECTION 4

DISKO ELM OUTPUT DATA AND THE SMART DATA MANAGER

Some 21 classified documents with output information regarding the Disko Elm UGT were scanned by JAYCOR and imported into the SDM. One of these DNA-furnished documents was not of sufficient quality to be included in final linked form in the SDM data base. These documents describe a variety of neutron, gamma-ray, and x-ray measurements and predictions. They consist of preliminary and final reports with time-dependent and time-integrated results: measurements, predictions, and consensus summaries. Figure 3 shows how output information can be stored and accessed in the SDM for a variety of events, devices, and types of radiation. For the specific application described here, only a single event (the Disko Elm UGT) has been considered.

Twenty classified documents were included in the final data base and required 13,725,474 bytes of memory. There are 158 pages of text, 131 tables, and 176 figures in the 20 documents. There are also 39 files of plot data in the one other document which were scanned but not incorporated into the final data base. There were also 270 columns of ASCII data created for plotting applications with MATLAB. The scanned plots are all in picture or TIFF form. Each of these files, even when compressed, typically requires about 50,000 bytes of memory. There are about 100 tables which are also in TIFF form and which require a significant amount of memory.

It took approximately 126 hours to scan the data base and 179 hours to edit and link the resulting data files. The most difficult documents to scan include multipage, small font, portrait-oriented tables, and tables which have columns of unequal lengths. TIFF files that must be rotated also require more steps and more time. Greek characters and symbols are not read correctly by Guide and need to be edited separately.

About 30 tables were converted to ASCII form; columns of data were stripped from these tables and placed in MATLAB files to provide a plotting capability. Additional tables were

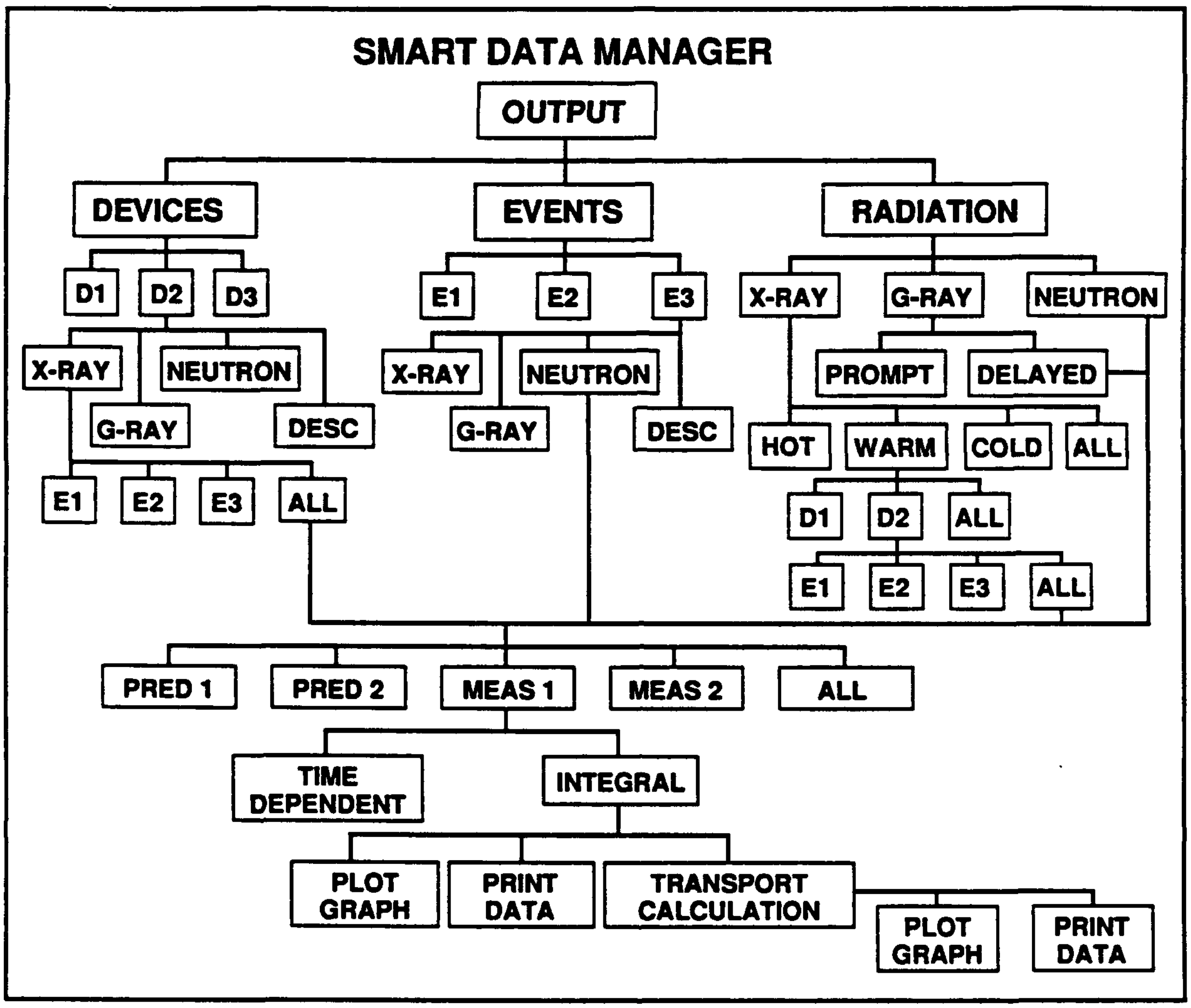


Figure 3. Typical tree structure for storing and retrieving archival output information.

converted to save memory, since tables in ASCII form require an order of magnitude less storage than TIFF tables. For this application, approximately 15 megabytes of disk storage are devoted to TIFF, PCX, GUIDE, and ASCII files. The three main programs in the SDM (WINDOWS, GUIDE, and MATLAB) occupy a total of about 9 megabytes of disk memory, for a total of 24 megabytes required. Various other packages, such as SCANGAL, OMNIPAGE, PAINTBRUSH, WORD PERFECT, NORTON COMMANDER, etc., which are used for scanning, interpreting, and manipulating data and text files can add another 10 megabytes of disk memory requirements.

The SDM was developed to run within the WINDOWS 3.0 graphical environment; this means that the SDM can be accessed by starting WINDOWS and simply clicking the mouse on an appropriate icon. For the Disko Elm output application, the user clicks on the "DNA OUTPUT" icon and Figure 4 appears. Figure 4 depicts the SDM window for the Disko Elm output application. By clicking the mouse on one of the live buttons in Figure 4, new windows are opened allowing the user to access and manipulate the stored information. For example, by clicking on the "OPEN DOCUMENTS" expansion button in Figure 4, the user accesses the window shown in Figure 5. This figure shows short titles of the 21 Disko Elm documents which are stored in the SDM. Each of these documents can be opened via the associated expansion button. (The expansion buttons are the circles with imbedded plus signs, and they are "live" between the left and right square brackets, see Figure 2.) Figure 5 also displays the next level of information for DE-16 when the expansion button for that particular document is activated. We see that the user may then open up the complete DE-16 document, look at references or distribution lists, or open up lists of plots and tables. Moreover, by clicking on one of the reference buttons (the open arrows shown in Figure 5), the user may view an individual table or plot of interest. The entire contents of DE-16 would be available to the user upon activation of the "COMPLETE DOCUMENT" reference button.

[* OUTPUT DOCUMENTS *]

[⊕ OPEN DOCUMENTS ⊕]

[⇨ *MAKE NEW PLOTS* ⇨]

[⇨ *HELP MENUS* ⇨]

Figure 4. Window taken from the SDM screen showing options for output data.

As indicated above, a number of tables within the Disko Elm document set have been converted to ASCII files from the scanned picture or TIFF data and the individual columns of data have been stored in MATLAB files. A plot of all data within one of these tables

[*OUTPUT DOCUMENTS*]

[⊕OPEN DOCUMENTS⊕]

[⊕DE-1	DISKO ELM Sieve Hole Closure Calculations, Len Reed, S3 Memo SSS-COTM-89-10542, 2 June, 1989.⊕]
[⊕DE-2	Calculated Gamma Dose Rates and Total Prompt Gamma Dose for DISKO ELM and Vacuum Scatterer Experiments⊕]
[⊕DE-3	DISKO ELM Preliminary Output Diagnostic Results⊕].
[⊕DE-4	DISKO ELM Preliminary Output Diagnostic Results⊕].
[⊕DE-5	DISKO ELM Time Dependent X-ray Spectra⊕].
[⊕DE-6	DISKO ELM Preliminary Low Energy X-ray Spectrum⊕].
[⊕DE-7	DISKO ELM Final Output Diagnostic Results⊕].
[⊕DE-8	Diagnostic Measurements on Project DISKO ELM Preliminary Results Report. ⊕]
[⊕DE-9	Preliminary Results Report Underground Test Output Diagnostics on DISKO ELM⊕].
[⊕DE-10	DISKO ELM Radiation Diagnostic Preliminary Results Report⊕].
[⊕DE-11	Final DISKO ELM Results in the Navy Vacuum Scatterer Results Report.

⊕]

[⊕DE-12	Prediction of the Disko Elm X-Ray Spectrum, A.J. Spero (2/13/89)⊕].
[⊕DE-13	Time Dependent X-Ray Output

Figure 5. Window taken from the SDM showing documents stored.

Calculations for Disko Elm (4/18/89)

⊕].

[⊕DE-14 Preliminary Results UGT Output Diagnostics on Disko Elm, R.I. Miller, D.N. Arion, and K.R.Sites.

⊕]

[⊕DE-15 Disko Elm Preliminary Results Tabulation, LMSC (10/11/89)⊕].

[⊕DE-16 Disko Elm Preliminary Output Diagnostic Results⊕]

- [⇨COMPLETE DOCUMENT⇨]
- [⇨REFERENCES⇨]
- [⇨DISTRIBUTION LIST⇨]
- [⊕PLOTS⊕]
  - [⇨FIGURE 1⇨]
  - [⇨FIGURE 2⇨]
  - [⇨FIGURE 3⇨]
  - [⇨FIGURE 4⇨]
  - □]
- [⊕TABLES⊕]
  - [⇨ TABLE 1⇨]
  - [⇨TABLE 2⇨]
  - [⇨TABLE 3⇨]
  - [⇨TABLE 4 ⇨]
  - [⇨TABLE 5⇨]
  - [⇨TABLE 6⇨]□]□].

[⊕DE-17 D+30 Results, UGT Output Diagnostics on Disko Elm⊕].

[⊕DE-18 Disko Elm Derivatives⊕].

[⊕DE-19 Gamma Radiation Measurements Disko Elm D+60 Preliminary Results Meeting ⊕]

[⊕DE-20 Disko Elm D+60 Radiation Diagnostic Results by Sandia Labs.⊕]

[⊕DE-21 Report to DNA on LLNL Diagnostics of Disko Elm Event. ⊕]

□]

Figure 5. Window taken from the SDM showing documents stored (Continued).

can be produced by clicking the mouse on a button within the table. Programming modifications of the hypertext and graphics routines necessitate only a single "run" command to generate such a plot. A typical plot (based upon fictitious data) is shown in Figure 6. Individual curves or subgroups of selected curves can also be easily generated by the user. In addition, titles of plots, labels for x and y axes, and grids are automatically generated. X and y coordinates for curves can also be retrieved by clicking the mouse at various locations on the screen to obtain the numerical values at these positions. All titles and labels can be readily modified by the user, and zooming and curve labelling capabilities are available.

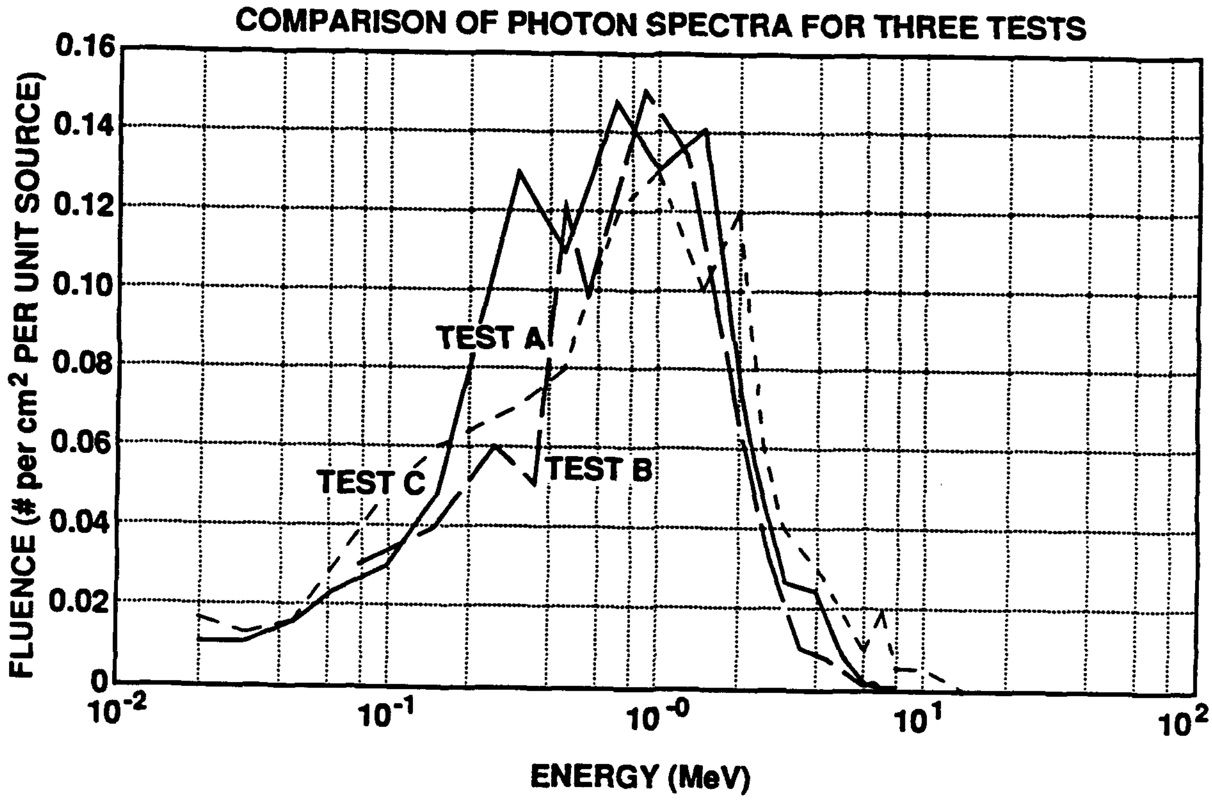


Figure 6. A typical plot which was directly obtained from a table of fictitious data.

Two other plotting schemes were developed to allow the user to generate plots from either a list of data files or from a tree-like index. (Approximately 270 columns of tabular data are in ASCII form.) These schemes are capable of plotting an individual curve or multiple curves, along with titles, labels, grids, etc.. By clicking the mouse on the "MAKE NEW

PLOTS" reference button of Figure 4, we obtain the SDM window of Figure 7. The activation of reference buttons for plotting files and the "OPEN PLOT TREE DATA" button leads to the SDM window of Figure 8. Figure 8 shows the first level of the plot tree index for the Disko Elm data. Various expansion buttons provide access to more detailed information and eventually enable the user to plot selected predictions and/or measurements.

A new and more comprehensive tree index was recently proposed by DNA personnel to better define and locate specific data sets. This index is displayed in Figure 9 and it has been partially implemented in the SDM.

Help menus have been incorporated into the SDM to assist the user with regard to: running the DNA output application, employing the WINDOWS 3.0, GUIDE 3.0, and MATLAB programs, scanning documents, linking files, and plotting data. Figure 10 is the SDM window which identifies the available help menus. To obtain descriptions and instructions, the user clicks on the menu of interest. Figure 11 is a plot help menu as it appears on the screen, along with a few of the MATLAB plotting commands.

Figure 12 displays a help menu for options in the GUIDE hypertext program. Activation of the index element of Figure 12 leads to the window of Figure 13, which is comprised of a "live" set of index buttons for GUIDE topics. Figure 14 shows GUIDE topics starting with the letter "R." These topics, in turn, are "live" and can be clicked on to get detailed instructions.

Similar help menus for the WINDOWS 3.0 program are available in the SDM.

[⇨⇨] [➡*1. OPEN PLOTTING FILES*➡]

[⇨*HELP MENUS*⇨]

[⊕PLOT HELP⊕]

```
x1=0; x2=0; x3=0; x4=0 ;x5=0; x6=0; x7=0; x8=0; x9=0; x10=0;
y1=0; y2=0;y3=0; y4=0; y5=0; y6=0; y7=0;y8=0; y9=0; y10=0
x2=x1;x3=x1; x4=x1; x5=x1; x6=x1; x7=x1 ;x8=x1; x9=x1; x10=x1
load f:\matlab\matlabe.mat
plot(x1,y1,x2,y2,x3,y3,x4,y4,x5,y5,x6,y6,x7,y7,x8,y8,x9,y9,x10,y10)

semilogx        semilogy        loglog
plot      gtext(' curve label')
xlabel('        ')
ylabel('        ')
title('             ')
grid
[x,y]=ginput
```

[⊕OPEN PLOT TREE DATA⊕]

[⊕LIST OF AVAILABLE PLOTS ⊕]

Figure 7. An SDM window used to open individual data for plotting.

[⊕OPEN PLOT TREE DATA⊕]

DISKO ELM PLOT DATA

[⊕XRAYS⊕]

[⊕SIEVED DATA⊕]

[⊕CORRECTION FACTORS FOR SIEVED DATA⊕]

[⊕CALCULATED INTENSITIES (1/keV)⊕] ❑]

[⊕PRELIMINARY SPECTRA⊕]

[⊕NORMALIZED DIFFERENTIAL EXPERIMENTAL DATA FINAL RESULTS ⊕]

[⊕CALCULATED SPECTRA⊕]

[⊕ESCAPE RATES⊕]

[⊕FINAL RESULTS⊕]

❑]

[⊕GAMMA RAYS⊕]

[⊕GAMMA RAY SPECTRUM⊕]

[⊕GAMMA DOT ⊕] ❑]

[⊕NEUTRONS⊕]

[⊕PREDICTED DIFFERENTIAL NEUTRON SPECTRUM⊕]

[⊕MEASURED DIFFERENTIAL NEUTRON SPECTRUM⊕]

[⊕PREDICTED NORMALIZED NEUTRON SPECTRUM⊕]

[⊕MEASURED NORMALIZED NEUTRON SPECTRUM⊕]

Figure 8. Window showing tree index structure for plotting data.

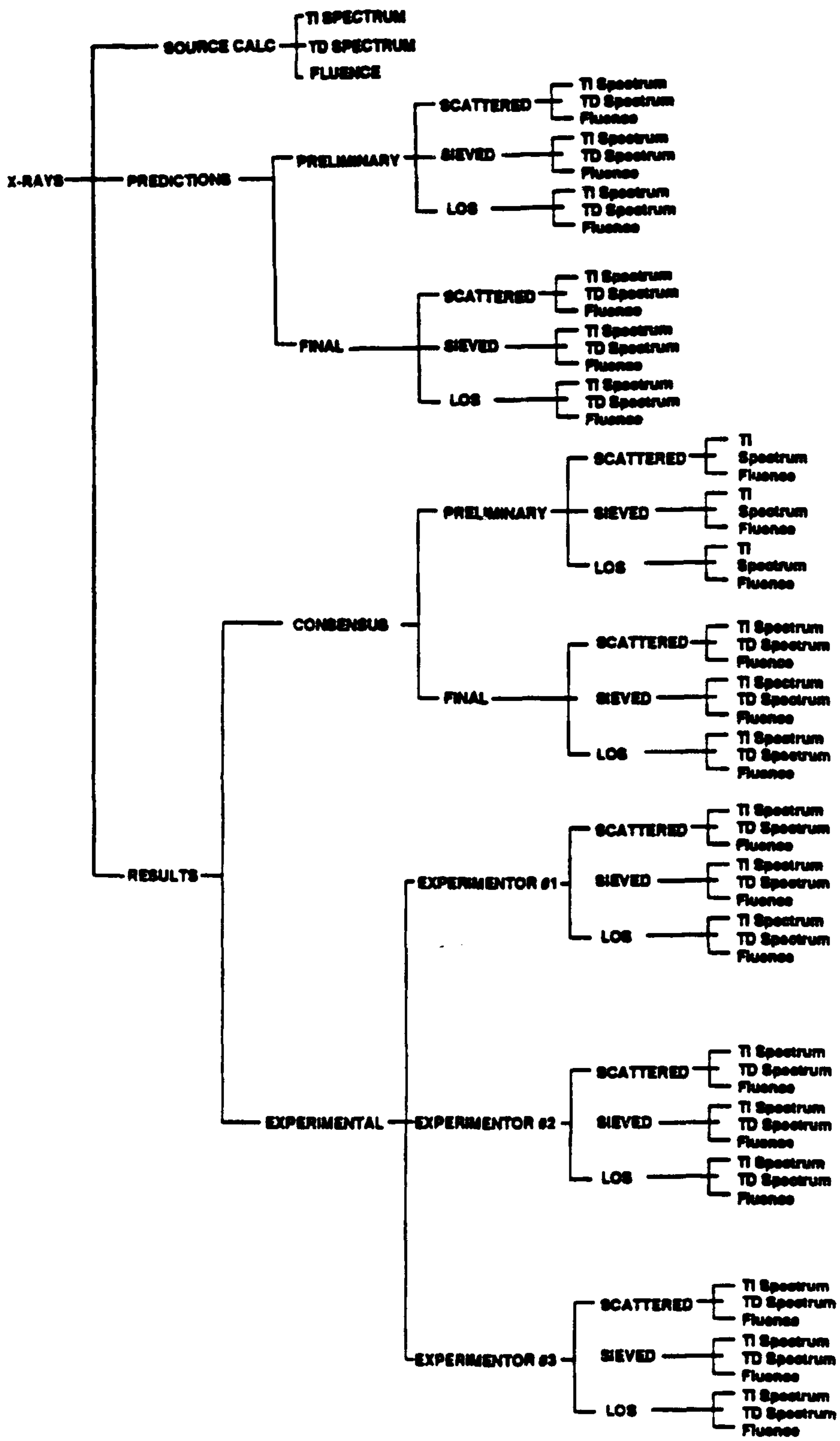


Figure 9. New tree structure for Disko Elm plots.

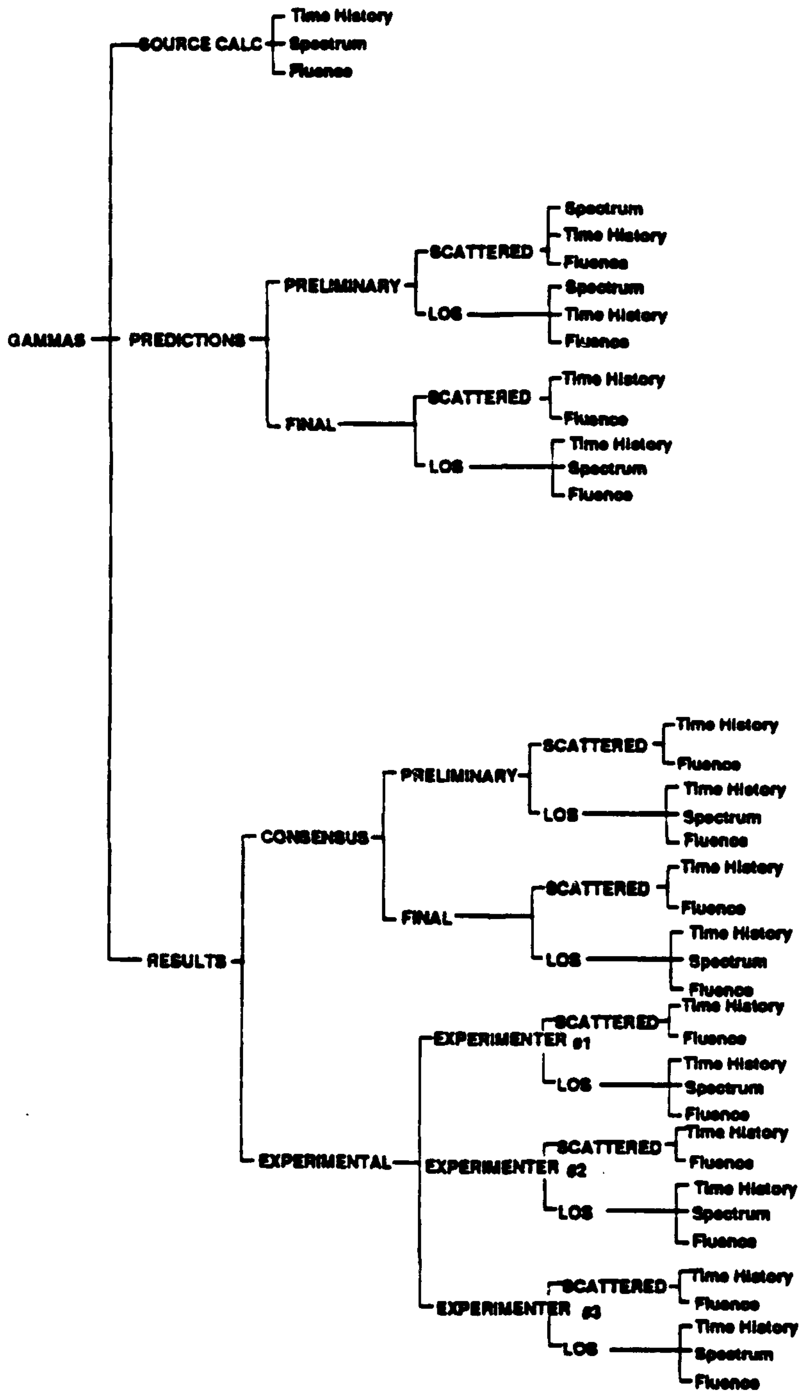


Figure 9. New tree structure for Disko Elm plots (Continued).

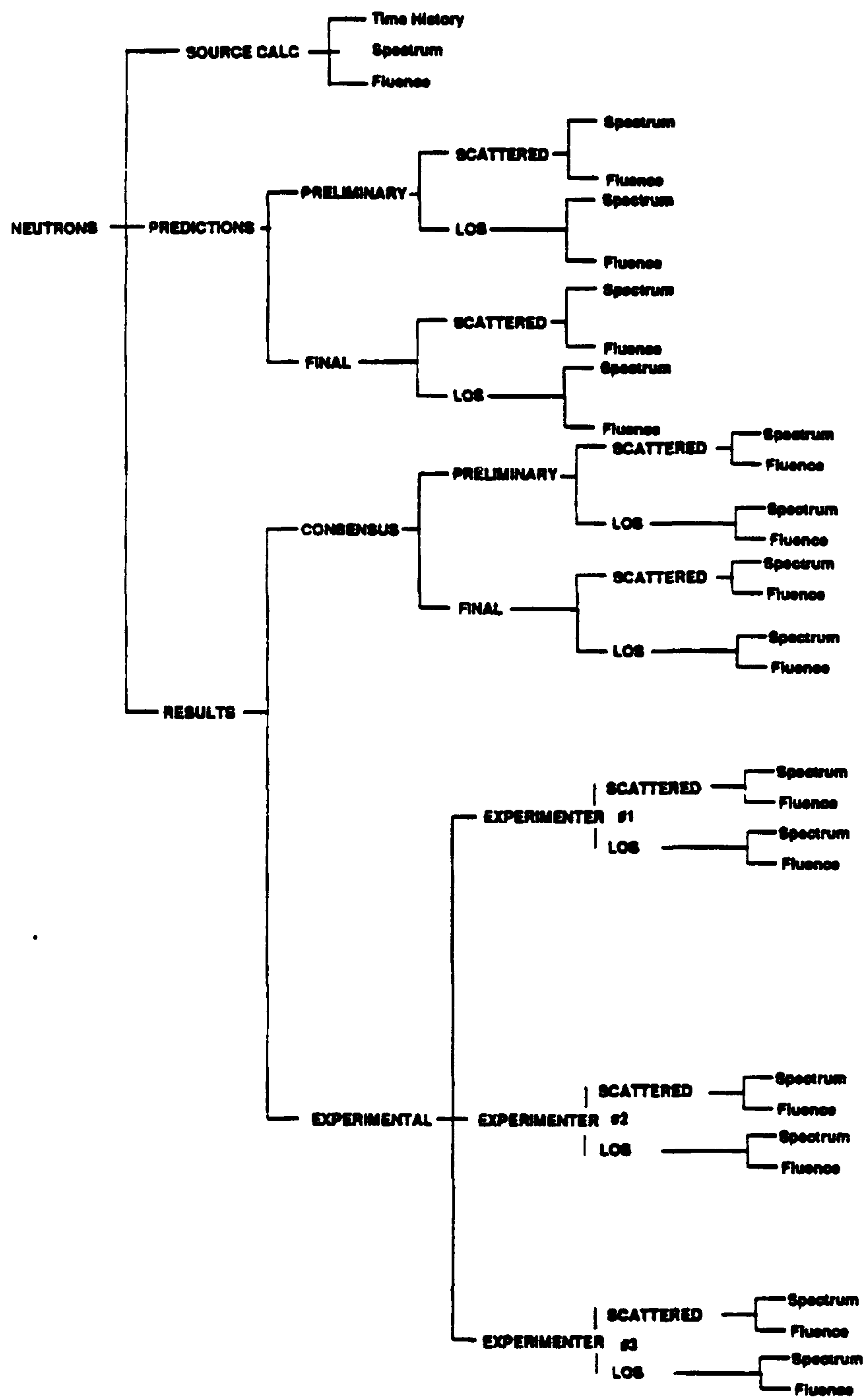


Figure 9. New tree structure for Disko Elm plots (Continued).

[⇨⇨][⇨⇨][⇨⇨][⊕**HELP MENUS** ⊕]

[⇨***RUNNING DNA OUTPUT***⇨]

[➡ ***WINDOWS 3.0 HELP***➡]

[➡ ***GUIDE HYPERTEXT HELP***➡]

[⇨***LINKING FILES IN GUIDE***⇨]

[⊕**MATLAB HELP** ⊕]

[⇨***PLOT LEGEND***⇨]

[➡***PLOTTING INSTRUCTIONS***➡]

[➡ ***GENERAL HELP***➡] (TYPE "help" AND RETURN WHEN MATLAB IS OPENED} □]

[⊕**OUTPUT GRAPHICS DIRECTIONS** ⊕]

[⇨***GRAPHICS DIRECTLY FROM ASCII TABLES***⇨]

[⇨***GRAPHICS FROM A LIST OF FILES OR FROM A TREE INDEX OF FILES***⇨] □]

[⊕**SCANNING HELP** ⊕]

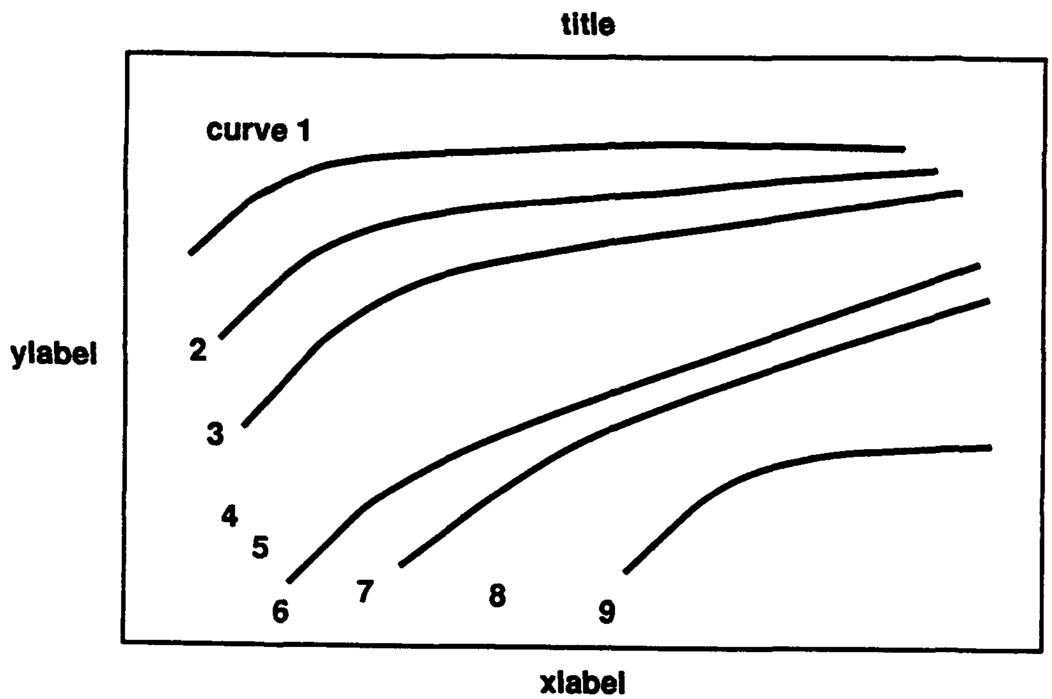
[⇨***SCANNING DATA***⇨]

[⇨ ***SCANNING IMAGES***⇨]

[⇨ ***ROTATING IMAGES***⇨] □]

□]

Figure 10. Smart Data Manager help menus taken from a screen window.



Use gtext(' name') - to place "name" on the plot with mouse
Use grid - to put a grid on the plot
Use [x,y]=ginput - to pull x-y coordinates off the plot with a mouse
xlabel(' name ') - places "name" on the x-axis
ylabel(' name ') - places "name" on the y-axis
title(' name ') - places "name" on the top of the plot

Figure 11. Plot help menu showing legend and MATLAB commands.

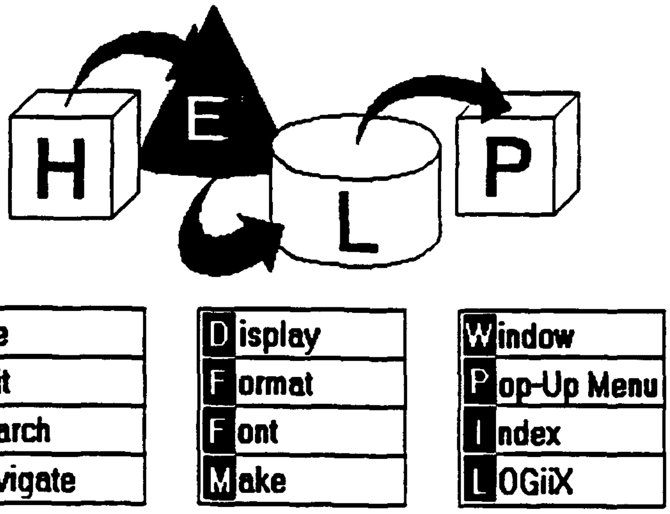


Figure 12. Overall help menu for the GUIDE program.

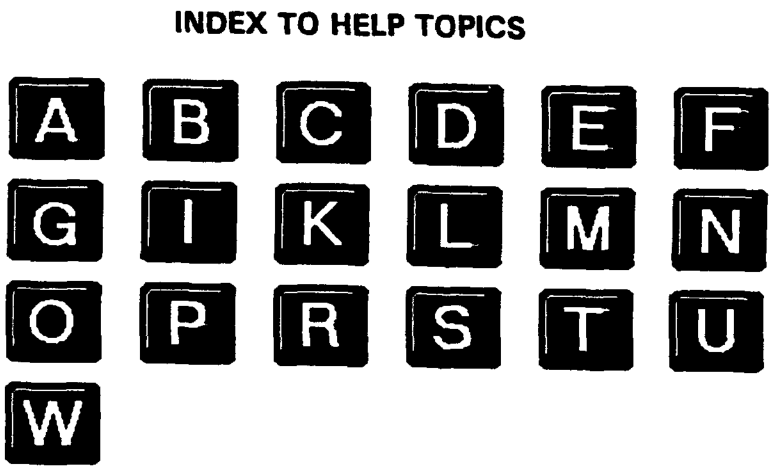


Figure 13. Help index for the GUIDE program.

INDEX TO HELP TOPICS

R

Readonly
- Open
- Save As

Reference Button
Reference Point
REL - Relative Measurement
Remain
Remote Link
Rename
Restore Size
Revert Content
Revert Window
Rich Text Format (RTF)
- Place
- Save As...

Figure 14. Help menu for GUIDE topics starting with letter "R."

SECTION 5

LESSONS LEARNED

A number of lessons have been learned from this initial project. It is clear that the SDM can be an important and attractive complement to a library of classified paper documents. The SDM provides a user-friendly approach to the storage, retrieval, and manipulation of vast amounts of technical information. Tables of data can be plotted, information can be extracted from such plots, and graphical comparisons of predictions with measurements can be made with relative ease using the SDM.

Paper documents which are in good condition can be scanned in a straightforward manner. Conversion of TIFF or picture files of text and tabular data can also be converted to ASCII form quite readily. However, data which are already available on diskettes can be even more efficiently utilized, since less effort is required compared to checking converted data as discussed above.

Some tabular data sets in paper documents are more difficult to scan and convert to ASCII form than others. These include tables with superscripts for exponents, very long tables spanning a number of pages, tables with extremely small fonts (since photocopies tend to have "filled-in" characters) and, especially, tables of poor quality. Most of the DNA-furnished data sets were of good quality.

Scanned plots and tables in the computer's memory can easily be printed out as hardcopies. In general, the quality of the hardcopy exceeds that of the screen image and is very close to the original plot or table in detail. However, one of the DNA-furnished documents (with the designator DE-21 in Section 4) was of poor quality; it was scanned but no attempt was made to create ASCII files and link them.

Nearly all of the textual material in the Disko Elm documents was scanned and converted into ASCII files with an accuracy rate of about 99 percent. Equations are more difficult to convert to text formats, but TIFF or picture files for the equations can be interspersed with converted text to achieve adequate computer documents. Also, note that scanned "landscape" figures can be rotated into more readable "portrait" forms using "paint" programs.

The SDM is a modular system so that a new graphics package can be introduced to replace MATLAB. EXCEL 3.0, for example, could be used to replace MATLAB as the primary SDM graphics module. MATLAB is a robust mathematics program including signal processing and statistical and engineering tools, in addition to graphical capabilities. It was selected for the SDM since no satisfactory WINDOWS-compatible application graphics software was initially available.

Because the MATLAB software was not designed to operate within the WINDOWS environment, the SDM must leave the WINDOWS environment to generate plots. Plotting 10 overlaid curves from a table may take 30 to 35 seconds on the 33-MHz 386 PC. About half of this time is spent in leaving the WINDOWS environment and in accessing MATLAB files via DOS. The employment of WINDOWS-compatible plotting programs could result in improved graphics computation times.

The QEMM 5.1 memory manager software has been used to more efficiently deal with programs which compete for memory within the WINDOWS environment. The 386 PC in Albuquerque had some memory allocation problems initially, but has performed quite well since the installation of the memory manager and introduction of a new consistent set of memory chips. None of the 386 PC memory and addressing problems were ever experienced on the 286 PC in San Diego.

WINDOWS 3.0 is a new program (introduced in May 1990) and more and more software is being updated to work within this operating environment. Even GUIDE 3.0, which was designed to run with WINDOWS 3.0, has recently been improved to deal with WINDOWS-related bugs. New drivers for the SYQUEST hard disk units have been introduced to treat other interface problems, and it is anticipated that many other programs will be updated to interface more smoothly with WINDOWS 3.0. New SYQUEST drivers specifically designed for WINDOWS 3.0 are also anticipated in the near future. In the meantime, there may continue to be minor compatibility problems with WINDOWS 3.0, but most of these are insignificant in comparison to the great merits of this operating environment.

SECTION 6

RECOMMENDATIONS

6.1 IMPROVED PLOTTING CAPABILITY.

The existing first-generation SDM was built by scanning a number of previously prepared UGT reports. Textual information and scanned graphics were organized using the GUIDE hypertext program, while interactive plotting and data manipulation was achieved using the MATLAB program.

Note that the GUIDE hypertext shell runs under the WINDOWS 3.0 extensions to the MS-DOS operating system. WINDOWS 3.0 provides a graphical user interface with all the easy-to-use features that make such a system appealing. Examples include pull-down or pop-up menus, 'point-and-click' methods for choosing items with a mouse, a 'clipboard' for 'cutting and pasting' both text and graphics to be shared between different applications, and extensive user control of fonts and formatting. The GUIDE program makes use of many of these features.

Unfortunately, MATLAB is not a WINDOWS-based program. Although it is a very powerful mathematical analysis program, it does not interface very well with the WINDOWS 3.0 system. It can be launched from GUIDE, and it does create interactive plots. However, operation of the SDM would be much smoother if the plotting package was designed to run under WINDOWS and to take advantage of the many user-interface features provided by WINDOWS. [Note that most older programs for standard DOS computers can be run under WINDOWS, but they do not have all the features available to programs written especially for WINDOWS. In these discussions, the term "WINDOWS-based program" is used to refer to software deliberately designed to run under the WINDOWS environment and to use various WINDOWS features.] For example, in a WINDOWS-based plotting program, graphs would appear in a separate window on the screen (which could be shown at the same time as the GUIDE window which contains the tabular data), individual plots

could be selected and manipulated by clicking the mouse and using pull-down menus, and copies of graphs could be passed to other WINDOWS-based programs (e.g., word processors) for printing or further manipulation. One would have the advantage that operation of the plotting program would be very similar to GUIDE or any other WINDOWS-based program.

Another advantage of using WINDOWS-based programs for plotting (or for providing other SDM features) is that WINDOWS includes some powerful capabilities for integrating different programs. Data can be interactively shared between various WINDOWS programs using 'Dynamic Data Exchange (DDE)', an information sharing protocol built into WINDOWS. Similarly, one can add new functionality to WINDOWS programs using 'Dynamic Link Libraries (DLL)'. These features mean that special capabilities (not even envisioned at the start of an effort) can be added as needs arise (thus making the SDM even "smarter").

In particular, it is recommended that some effort be spent looking into various WINDOWS-based programs for scientific plotting. JAYCOR's previous experience in building the current SDM has shown that one very important type of data might be called the "set of xy pairs of numbers." These sets of pairs of numbers might be displayed as columns in a table or as a curve in an xy graph. Such sets represent some functional relationship (such as the value of a parameter as a function of time), and the ability to easily manipulate and display such relationships is an important aspect of any smart data base manager.

Ideally, there are a large number of operations that the SDM would be able to do with the various collections of xy pairs of numbers described above. The most obvious, of course, is to be able to plot the pairs as a curve on an xy graph, and to be able to manipulate the scales, size, and format of that graph. One should be able to easily select the pairs (preferably by simply 'pointing' to either a data table or the actual curve on a graph) from different sources and then easily overlay the data on a single plot for comparison purposes. One should also note that the data are not simply raw numbers; in most cases, the numbers

will have some unit (e.g., time in seconds, field strength in volts/m) and each set of numbers will also have some identifying textual description (i.e., a label). This information needs to move with the numbers as comparisons and manipulations are carried out. A "smart" data base manager would be smart enough to automatically carry out such things as unit conversions when data sets with dissimilar units are compared.

Many desirable characteristics of a WINDOWS-based plotting program are already beginning to appear in commercial software. A good example is the new version of Microsoft's spreadsheet, EXCEL 3.0. [Examples of EXCEL 3.0 windows are shown in Figure 15.] Data can be easily entered or transferred to and from other programs (using files, the WINDOWS clipboard, or DDE techniques). Plots can be created and formats extensively varied and modified. [The user has complete control over scales, labels, axes, grids, colors, fonts, etc.] Curves can be selected by simply using a mouse (or the keyboard) and selected curves moved to other plots and compared to other sets of curves. EXCEL 3.0 even includes hypertext features like programmable 'buttons,' and an outlining feature for creating 'tree structures' of related information. EXCEL is a general purpose spreadsheet program, but it is not optimized for scientific and technical data. It can be highly customized, however, using the built-in macro language, and complex extensions can be added using various features of the WINDOWS operating system. For instance, although units are not normally attached to sets of number pairs, one can customize EXCEL to keep track of units and convert to other units when necessary.

Another possibility is a custom-written WINDOWS-based plotting program. Such a program has the advantage that its features can be customized exactly as specified. JAYCOR already has one such custom written program operational, and it could be polished up for use with the SDM.

It should also be noted that GUIDE is not the only hypertext shell now available for use under WINDOWS 3.0. As previously mentioned, the new version of the EXCEL spreadsheet now has some hypertext capabilities, and other programs like *ToolBook*,

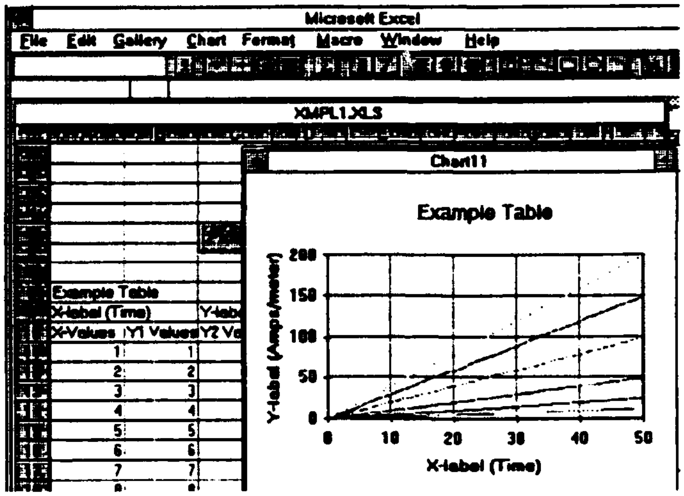


Figure 15. EXCEL 3.0 screen showing graph and tabular data.

KnowledgePro, and *Plus* are now also available. Each of these programs has some advantages (and disadvantages). It would be useful to have some comparison of capabilities in order to help guide future work.

6.2 MASSIVE DATA STORAGE.

If large numbers of reports are to be scanned and linked, the SDM will have to be augmented with massive data storage devices. The need for large amounts of storage is particularly great when the data includes numerous scanned graphical images. An uncompressed scanned image may generate a file of a megabyte or more. Data compression techniques may reduce the storage requirement by a factor of 10 to 20, but it is clear that a complete set of UGT documentation could necessitate more space than is available on the 44-megabyte Syquest drives used in the current SDM at JAYCOR.

Conventional magnetic hard disk drives are available with capacities of 200 to 500 megabytes, and there are some new magnetic storage systems with large-capacity removable media. Various optical storage devices now exist with removable media and storage capabilities of thousands of megabytes. Such systems include WORM (write-once, read

many) media, erasable optical systems (magneto-optical), and CD-ROM (read only). A comparison of storage capabilities and access time for various storage techniques is shown in the Figure 16.

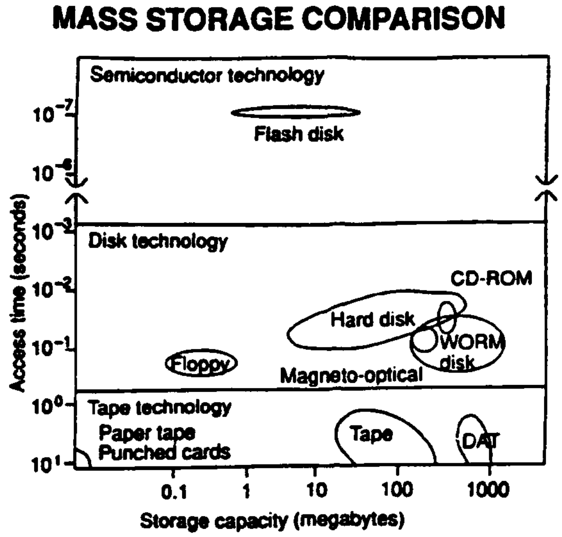


Figure 16. Mass storage comparisons (from BYTE magazine, Nov. 1990).

One can see that the optical storage techniques offer the capability to handle large amounts of data, but have access times typically slower than hard disks.

One reported problem with WORM drives is the lack of a standard file format. Drives from different manufacturers typically will not read disks written on a machine from a different manufacturer, and the interfaces to DOS are often different. Some of these standardization problems are reportedly being worked out, but some investigation and actual hardware trials will undoubtedly be needed to insure that any system would work with WINDOWS 3.0 and the current SDM configuration. One recommended effort is thus the investigation of various alternatives for massive data storage by the SDM.

6.3 DIGITIZATION.

Quantitative information may only be available in the form of a graph, rather than as a table of numerical data. Scanning such a graph turns it into bits where, for example, a one

may represent a pixel that is black and a zero may represent a white (i.e., background) pixel. Such images are called 'bitmaps' or 'raster-scanned' images. For comparison purposes, one would really like to have the curves in such an image turned into xy pairs of numbers. These numbers can then be manipulated to produce plots useful for direct comparisons. Ideally, one would like to have an automatic digitization program that would turn bitmaps of graphs into tables of numbers (just as OCR programs turn bitmaps into text by recognizing the letters of the alphabet).

Several such automatic digitization programs do reportedly exist (at least on Macintosh computers), but JAYCOR has no direct experience with their capabilities. Undoubtedly, they have problems similar to those of OCR systems (bitmaps with 'noise' from poor original copies, errors due to the presence of grid lines, etc.). Such programs should be investigated, however, to see how useful they may be. [Note that a related capability is now included in a number of graphics software packages. This is the ability to automatically 'trace' bitmapped images and turn them into object-oriented art; the process is called 'bitmap-to-vector' translation.]

A manual digitization scheme is more readily achievable and, in fact, a simple scheme has already been implemented in JAYCOR-produced electronic handbook material. One first displays the bitmapped image on the screen and then draws a transparent rectangle over the image. Screen pixel locations for the corners of the rectangle are noted, along with the corresponding floating point values of the graph. Next, one simply moves the mouse cursor to any point on the graph, clicks the mouse button, gets the cursor location, transforms it to the coordinate system of the graph, and displays the coordinates of the indicated point. The software for such a manual digitizer has already been written under the WINDOWS program "ToolBook." We recommend a more flexible implementation that would allow the digitization of a displayed image by any program which runs under WINDOWS. [This scheme would utilize the WINDOWS multitasking capability to run a second program with a transparent window; DDE and/or DLL techniques could then be employed to share information between the two programs.]

One limitation of this on-screen digitization approach is that its accuracy is determined by the number of screen pixels and the size of the displayed image. Since a VGA display has a resolution of 640 x 480 pixels, one can see that the accuracy is limited to one part out of a few hundred (i.e., a few tenths of a percent). For many purposes, this accuracy is quite sufficient, but if one tries to digitize a linear plot and then redraws the information on a log-scale graph, the accuracy of some of the points may be questionable.

6.4 NEW TREE INDEX FOR DISKO ELM PLOTS.

As previously indicated, a comprehensive tree-like data index has been proposed by the DNA personnel to help define and locate the various data sets of interest. The tree structure is shown in Figure 9. This structure has already been integrated into GUIDE, but the links to the actual data and graphs have yet to be established. A complete implementation of the tree will be a significant effort because not all of the necessary data is in the SDM, and because it is a nontrivial undertaking to insure that many data links are correctly implemented. [To be sure, a good part of the 'intelligence' of the SDM is in the links between related data.]

Tools other than GUIDE (which is primarily aimed at organizing textual documents) might also help organize and find tabular data or charts. The EXCEL 3.0 program includes an 'outlining' feature which makes it possible to easily put tabular or graphic information into a tree-like structure, and the user defines the level of detail to be displayed. Figure 17 shows an EXCEL 3.0 outline for the proposed x-ray data structure. The left side of the figure shows only three levels. By clicking on the symbols to the left, one can open the various levels to show further details. This is demonstrated on the right side of the figure where "Preliminary - Predictions" have been opened to show additional levels of the tree structure.

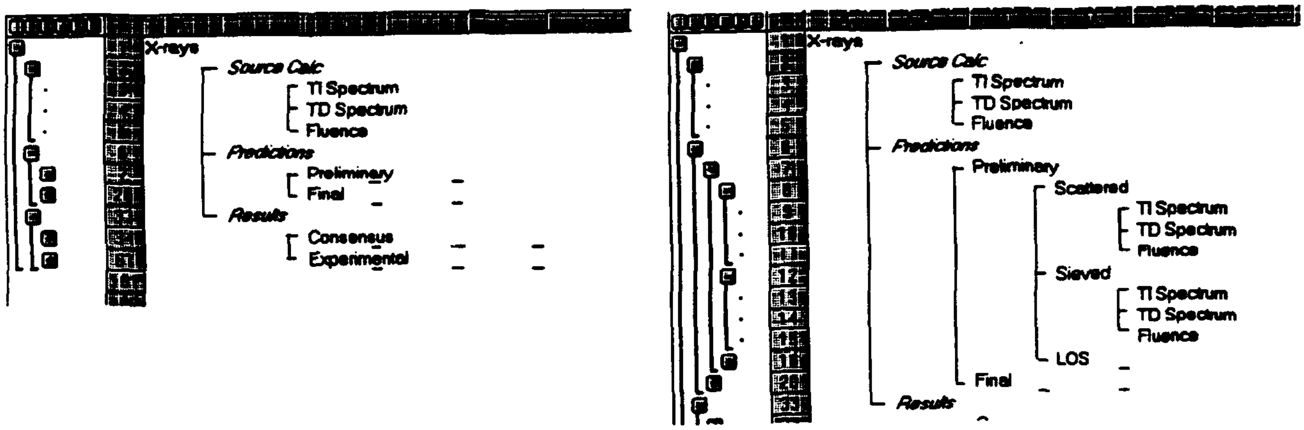


Figure 17. An EXCEL 3.0 outline for the organization of data.

6.5 COLOR OUTPUT.

WINDOWS 3.0 and GUIDE demonstrate that color computer displays can enhance the information content. In the present SDM documents, buttons of different colors indicate different levels of information or detail (i.e., different levels of a tree-type structure). Colors also distinguish different curves in a complex graph. Colors can thus help the viewer to understand complex data, in addition to just improving the visual appeal of a display.

There are now several manufacturers of color Postscript printers (notably Tektronix and QMS) with list prices under $10,000. These printers have resolutions of 300 dots per inch and can produce 16 million different colors! Output can be printed on paper or on transparencies. JAYCOR has two of these printers in its San Diego headquarters so that their compatibility with the SDM can be readily explored.

An additional benefit of relying on WINDOWS-based programs in the SDM is the fact that WINDOWS provides the printing environment. This means that the printer drivers for all the different output devices are part of WINDOWS. WINDOWS provides a Postscript driver, so printing color copies of WINDOWS screens should be quite easy. [The comment 'should be' is used because we have not actually tried to drive JAYCOR's color printers with the WINDOWS Postscript driver.] Also, the ability of WINDOWS-based programs to

share graphics via the 'clipboard' means that a graph prepared in one program (e.g., EXCEL) could be easily passed to a drawing or presentation program (e.g., POWERPOINT) for further enhancement or incorporation in a presentation.

A relatively inexpensive color output device is the color inkjet printer. Hewlett-Packard makes a system with 180 dots per inch resolution which sells for about $1,500. Slide-making devices that are directly attached to computers also exist, as do several types of color LCD panels that are driven by PCs and placed directly on an overhead projector for showing computer displays to a large audience. Lastly, one can prepare an unclassified color presentation on a PC and then employ a modem to send the slide information to a remote location. The color slides or transparencies are then prepared at this location and returned to the originator by overnight delivery service. JAYCOR has used this capability (which is a built-in feature of Microsoft's POWERPOINT software) to obtain color slides for several conferences.

DISTRIBUTION LIST

DNA-TR-91-134

DEPARTMENT OF DEFENSE

DEFENSE INTELLIGENCE AGENCY
ATTN: DB-6

DEFENSE NUCLEAR AGENCY
ATTN: DFRA
ATTN: DFSP
ATTN: DFTD
ATTN: IRMO E K HONG FONG
ATTN: TDTR
2 CYS ATTN: TITL

DEFENSE TECHNICAL INFORMATION CENTER
ATTN: DTIC/FDAB

FIELD COMMAND DEFENSE NUCLEAR AGENCY
ATTN: FCNV J LACOMB

FIELD COMMAND DEFENSE NUCLEAR AGENCY
2 CYS ATTN: FCTT W SUMMA
ATTN: JOE QUINTANA
2 CYS ATTN: TDTT LTCOL GAUDET
ATTN: TTST E RINEHART
ATTN: TTST R SHOUP
ATTN: TTTT RUSSELL B DANIEL
ATTN: TTUT LT COL NAEGELI

DEPARTMENT OF ENERGY

LAWRENCE LIVERMORE NATIONAL LAB
ATTN: L-221 B HUDSON
2 CYS ATTN: TECH LIBRARY

LOS ALAMOS NATIONAL LABORATORY
ATTN: C DIVISION R MCLEAN
2 CYS ATTN: REPORT LIBRARY
ATTN: R KIRBY

SANDIA NATIONAL LABORATORIES
ATTN: A CHABAI DIV 9311
ATTN: DIV 9313 DR S R DOLCE
ATTN: DIV 9321 W BOYER
ATTN: F BIGGS DIV 9312
ATTN: ORG 9310 J D PLIMPTON
2 CYS ATTN: TECH LIB 3141

DEPARTMENT OF DEFENSE CONTRACTORS

JAYCOR
2 CYS ATTN: D HIGGINS

JAYCOR
2 CYS ATTN: D SARGIS

JAYCOR
ATTN: CYRUS P KNOWLES

JAYCOR
2 CYS ATTN: B TRAUTLEIN

JAYCOR
ATTN: R POLL

JAYCOR
ATTN: J BOMBARDT

KAMAN SCIENCES CORP
ATTN: DASIAC

KAMAN SCIENCES CORPORATION
ATTN: DASIAC

R & D ASSOCIATES
ATTN: G IVY

R & D ASSOCIATES
ATTN: B KILLIAN
ATTN: L GERMAIN

S-CUBED
ATTN: S PEYTON

SCIENCE APPLICATIONS INTL CORP
ATTN: DR M MCKAY
ATTN: TECHNICAL REPORT SYSTEM

THE TITAN CORPORATION
ATTN: S SCHUSTER